

## REMARKS

*Generally*

To anticipate a claim, the reference must teach every element of the claim. MPEP 2131.01. In rejecting most claims in the application, the Office Action (OA) mischaracterizes the technology disclosed by in U.S. Patent Application Publication No. US 2002/0059161 to Wen-Syan Li *Supporting Web-Query Expansion Efficiently Using Multi-Granularity Indexing and Query Processing* (Li) to find claim elements not present in Li.

Each claim of the application is addressed to the use of document comparison using **Latent Semantic Indexing (LSI)**. The OA relies on Li [0016] (and to a lesser extent on Li [0015]) in combination with various other (uniformly misapplied) sections of Li to reject each claim. Importantly, **Li teaches away from the use of LSI** by including it in the Description of the Related Art and concluding that section with:

... the past work has failed to address the problem of efficient processing of queries when they undergo query expansion or of reducing the size of the indexes used to perform query expansion and processing.

Consequently, the OA fails to state a *prima facie* case of anticipation under 35 U.S.C. §102 by citing disclosure that **teaches against** the use of LSI in combination with other uniformly mischaracterized portions of Li.

Further, 37 C.F.R. 1.104 Nature of Examination states:

*(c ) Rejection of Claims ... (2) ... The pertinence of each reference, if not apparent, must be clearly explained ...*

The OA cites to Li over one hundred (100) times; but explains the pertinence of the reference in only sixteen (16) instances. In fifteen (15) of those sixteen (16) instances, the explanation is limited to an assertion that:

<words found in the claim > represent / reads on <words found in Li>

The extent and specificity of the following remarks is a strong indication that the pertinence of each reference used in the OA is not apparent (as required by 37 C.F.R. 1.104). Should the Examiner find the present remarks unpersuasive to allow each claim, the undersigned requests another non-final action with the pertinence of each reference explained.

In general, LI is insufficient as a 35 U.S.C. § 102 reference against the claimed invention.

*Regarding the Specification*

Regarding the title of the invention.

The OA requires that the application bear a title

that is clearly indicative of the invention to which the claims are directed.

The OA then suggests

Method for document comparison and selection using latent semantic content and term tuple.

Given that several of the claims are not directed to tuples, but all the claims are related to latent semantic indexing, the title of the specification has been changed to:

DOCUMENT COMPARISON AND SELECTION USING LATENT SEMATIC  
INDEXING

Regarding the arrangement of the specification.

The application, as originally submitted, was formatted for electronic filing. Technical difficulties prevented electronic filing. To facilitate examination, the specification has been reformatted to comply with current rules for paper applications. No changes, other than the title noted above, have been made to the content of the written description portion of the specification.

*Regarding Claim Objections*

As noted above, the application as originally submitted, was formatted for electronic filing. The two occurrences of electronic filing artifacts, i.e., \_Ref532030037 in Claim 14, and \_Ref532038902 in Claim 18 have been removed.

In addition, a typographical error in Claim 3 has been corrected to show that claim's proper dependence on Claim 2.

*Regarding the Claim Rejections Under 35 U.S.C. § 102 as Anticipated by LI.*

Regarding Claim 1 and claims dependent from it, LI does not teach every element of Claim 1.

The OA asserts:

LI discloses a method for representing the latent semantic content of a plurality of documents, each document containing a plurality of terms.

The OA cites LI [0015], [0016], [0035], Fig. 2(a) and 2(b), [0096], [0103], and [0035] as specific portions of LI anticipating the claimed invention. Other than the assertion above (directed to [0016]), the OA fails to explain the pertinence of each reference.

LI does not teach or enable the elements of the claim as arranged in the claim. Specifically, LI does not disclose the use of n-tuples as entries in a term x document matrix that is subject to singular value decomposition and dimensionality reduction as claimed.

LI does not disclose *n-tuples* as claimed. LI [0035] discloses:

The present approach takes the indices for semantically similar and syntactically related words ... and further reduces the index size by merging some entries (tuples) to an entry at a higher level of granularity. These tuples are shown throughout LI as being of the form {<word>, <related word 1>, ... , <related word n>}.

Throughout LI (e.g., Fig. 2, Fig. 3, Fig. 4, [0044]-[0046], [0048]-[0053], [0091]) tuple relates to semantically similar and syntactically related words without considering their order of

appearance in a document. This is contrary to the present application where *n-tuples* are preferably formed from contiguous words, e.g.,

*[0037] One method of identifying n-tuples is to consider n contiguous words in a document as an n-tuple.*

LI does not disclose a two dimensional matrix as claimed. LI Fig. 2 discloses a **table** where each row relates to a term, and each table entry relates to a **list** of documents containing the term. Separately, LI [0096] discloses

... a two dimensional ranking graph... for a query with two words.

Even though this ranking graph is represented as a matrix, the rows do not correspond to terms, the columns do not correspond to documents, and the matrix elements do not correspond to the frequency of a given term in a given document - all this as claimed. LI [0103] contains no disclosure whatsoever related to matrix representation.

LI [0035], discussed above, discloses the use of an index where each index entry may be more than one term or document ID. Claim 1 of the application calls for entries to be

*related to the number of occurrences*

of a term - dramatically different than an index table as disclosed in LI [0035]. In addition, the OA mischaracterizes LI's word list as a matrix row corresponding to a term. The rows of the tables in LI Fig. 2 correspond to: documents in Fig. 2(a); and single terms in Fig. 2(b). The table entries in LI Fig. 2a are each a word list (emphasis added); whereas the matrix elements as claimed are related to the number of occurrences of a term in a document.

While LI [0015] and [0016] identify the use of Latent Semantic Indexing (LSI) generally, LI does not disclose a matrix representing the collection as claimed. The referenced paragraphs do not contain disclosure of "n-tuples" indexed into an LSI space. Importantly, LI **teaches away** from the use of LSI by including it in the Description of the Related Art and concluding that section with

... the past work has failed to address the problem of efficient processing of queries when they undergo query expansion or of reducing the size of the indexes used to perform query expansion and processing.

No portion of LI teaches or suggests creation of an LSI space using n-tuples as claimed.

As a consequence, the invention as claimed in Claim 1, and claims dependent from it, is patentably distinct from LI.

Regarding Claim 2, LI does not teach every element of the claim.

The OA asserts LI Claim 63, [0089], and [0035] without explanation of the pertinence of these references.

LI does not disclose *n-tuples* as claimed, either by composition or with respect to the use of a threshold for including the n-tuple in a term-by-document matrix. See comments directed to Claim 1 for rationale, e.g., with regard to LI [0035].

Further, LI [0089] discloses,

Turning now to a ranking scheme ... word representations at a coarser granularity are used for filtering out unrelated documents.

The OA provides no indication of how a scheme for ranking the results of a query relates to criteria for including a particular n-tuple in a term-by-document matrix. In fact, Claim 2 includes no mention of a query. The claim is directed to representing a document collection in an LSI space.

As a consequence, and in addition to the reasons cited with regard to Claim 1, the invention as claimed in Claim 2 is patentably distinct from LI.

Regarding Claim 3, LI does not teach every element of the claim.

The OA asserts LI Claim 10, Claim 25 without explanation of the pertinence of these references.

LI Claim 10 and LI Claim 25, while mentioning the use of thresholds, cannot be seen as anticipating the use of thresholds as claimed. As a consequence, and in addition to the reasons cited regarding Claim 2, the invention as claimed in Claims 3 is patentably distinct from LI.

Regarding Claim 4, LI does not teach every element of the claim.

The OA cites LI [0074] - [0080], and [0035] without explanation of the pertinence of these references.

Neither paragraph [0035] (its inapplicability described above) nor paragraphs [0074]-[0080] of LI discloses creating *n-tuple* terms from consecutive verbatim terms for use in creating an LSI space, as claimed. LI [0074]-[0080] are directed to processing words in the query, e.g.,

Furthermore there are m dictionary words and n non-dictionary words in the query Q. Given such a query the multi-granularity query expansion is performed as ... for each s1 ... replace S1, in Q ...

The invention as claimed creates n-tuples for the LSI space representing the collection of documents. There is no limitation claimed for query for query expansion.

As a consequence, and in addition to the reasons cited regarding Claim 1, the invention as claimed in Claims 4 is patentably distinct from LI.

Regarding Claim 5, LI does not teach every element of the claim.

The OA cites [0060], [0035], [0015], Fig. 2(a), Fig. 2(b), [0097], [0096], [0016], [0081], and [0035] without explanation of the pertinence of these references other than to assert:

"determining conceptual similarity" reads on "grouping dictionary words into semantic concepts"

with regard to [0060], and

"term" reads on "word list"

with regard to Fig. 2(a) and Fig. 2(b).

LI [0060] relates to:

the storage overhead of the on-line dictionary and the syntactic co-occurrence table that are needed to support basic query expansion <emphasis added>.

LI goes on to establish a variable  $f$  as

the compression factor <for storage> that is obtained by grouping words into semantic concepts <emphasis added>.

Storage overhead, as discussed in LI, is irrelevant to Claim 5. On a broader note, by its own definition LI groups words - LI doesn't determine conceptual similarity between documents.

LI does not disclose *n-tuples* as claimed. At the cited paragraph, LI discloses:

The present approach takes the indices for semantically similar and syntactically related words ... and further reduces the index size by merging some entries (tuples) to an entry at a higher level of granularity. During query processing the tuples with information at a higher level of granularity are used to retrieve relevant documents.

These tuples are shown throughout LI as being of the form {<word>, <related word>, ... , <related word  $n$ >}. Throughout LI (e.g., Fig. 2, Fig. 3, Fig. 4, [0044]-[0046], [0048]-[0053], [0091]) tuple relates to semantically similar and syntactically related words without considering their order of appearance in a document. This is contrary to the present application, *n-tuples* are formed from contiguous words, e.g.,

*[0037] One method of identifying n-tuples is to consider n contiguous words in a document as an n-tuple.*

LI does not disclose formation of a plurality of matrices, as claimed. The cited paragraph [0015] discloses the formation of a single  $M \times N$  matrix and is later decried as past work that has failed to address the problem.

LI does not disclose a two dimensional matrix as claimed. At the cited Figure, LI discloses a table where each row relates to a term, and each table entry relates to a list of documents containing the term. LI [0097] discloses a two dimensional ranking graph for a query with two words. Even though this ranking graph is represented as a matrix, the rows do not

correspond to terms, the columns do not correspond to documents, and the matrix elements do not correspond to the frequency of a given term in a given document - all this as claimed.

The OA mischaracterizes LI's word list as a matrix row corresponding to a term. The rows of the tables in LI Fig. 2 correspond to: documents in Fig. 2(a); and single terms in Fig. 2(b). The table entries in LI Fig. 2a are each a word list; whereas the matrix elements as claimed are related to the number of occurrences of a term in a document.

LI does not disclose matrix elements representing the number of occurrences of a term in a document, as claimed. LI [0096] discloses a ranking graph for the results of a two-word query. Claim 5 of the application calls for entries to be *related to the number of occurrences of a term* - dramatically different that the two-dimensional ranking graph disclosed in LI [0096].

While LI [0016] identifies the use of Latent Semantic Indexing (LSI) generally, the cited paragraph does not disclose a matrix representing the collection as claimed. The referenced paragraph, nor any other section of LI, does not contain disclosure of *n-tuples* indexed into an LSI space. Importantly, LI **teaches away** from the use of LSI by including it in the *Description of the Related Art* and concluding that section with

... the past work has failed to address the problem of efficient processing of queries when they undergo query expansion or of reducing the size of the indices used to perform query expansion and processing.

Further, LI does not teach nor suggest use of a plurality of LSI matrices, as claimed. Further, LI [0081] is wholly unrelated to singular value decomposition and the subsequent dimensionality reduction characteristic of LSI. LI [0081] describes the meaning of a compression factor *f* and how *f* indicates the utility of LI's invention for reducing storage space through a means other than LSI.

LI [0035], or any other section of LI, does not contain disclosure of "n-tuples" indexed into an LSI space. Importantly, LI teaches away from the use of LSI by including it in the *Description of the Related Art* and concluding that section with ... the past work has failed to address the problem of efficient processing of queries when they undergo query expansion or of



reducing the size of the indexes used to perform query expansion and processing. Further, LI does not teach nor suggest use of a plurality of LSI matrices.

LI [0057]-[0060] is disclosure directed to estimating storage overhead. Such disclosure is totally irrelevant to determining a composite similarity measure between a subject document and one or more reference documents. LI [0038] is disclosure directed to LI Fig. 3 and the idea of semantic similarity and syntactic relation between words. Neither LI [0038] or LI Fig. 3 even mention documents at all.

As a consequence, the invention as claimed in Claim 5 is patentably distinct from LI.

Regarding Claim 6, LI does not teach every element of the claim.

The OA cites LI [0035], and [0015] -- [0017], without explanation of the pertinence of these references.

In addition to the reasons cited regarding rejection of Claim 5, Claim 6 is patentable over LI for the following reasons. Earlier remarks note that LI does not anticipate the use of n-tuples in an LSI space - in LI [0035], [0015], [0016], [0017] or anywhere else - and that LI actually teaches away from use of an LSI space. Further, LI surely does not disclose the use of multiple LSI spaces, weighted or otherwise.

As a consequence, and in addition to the reasons cited regarding Claim 5, the invention as claimed in Claim 6 is patentably distinct from LI.

Regarding Claim 7, LI does not teach every element of the claim.

The OA cites [0016], [0017], [0018], Fig. 8, [0057] without explanation of the pertinence of these references other than to assert:

"1st", shows "verbatim entities" represented by "exact match."

With regard to Fig. 8, and

"stemming" reads on "grouping words by using word stemming"

with regard to [0057].

While LI [0016]-[0018] identifies the use of Latent Semantic Indexing (LSI) generally, the cited paragraph does not disclose a matrix representing the collection as claimed. The referenced paragraphs, nor any other section of LI, do not contain disclosure of stemming verbatim terms (those terms contained in the original documents before applying LSI) and using such stemmed terms in an LSI space by supplementing the verbatim terms. Importantly, LI teaches away from the use of LSI by including it in the Description of the Related Art and concluding that section with:

... the past work has failed to address the problem of efficient processing of queries when they undergo query expansion or of reducing the size of the indexes used to perform query expansion and processing.

LI [0057] mentions stemming, but not used as claimed here. LI [0018] discloses nothing regarding supplementing an LSI space with stemmed verbatim terms. The exact match of LI Fig. 8 does not relate to *verbatim entities* as disclosed in the application. The application refers to the original terms of the reference documents as *verbatim*, see for example

*[0029] In some embodiments, the present invention includes a method for representing the latent semantic content of a plurality of documents, where each document contains a plurality of verbatim terms.*

As a consequence, the invention as claimed in Claim 7 is patentably distinct from LI.

Regarding Claim 8, and Claims 10-13 dependent therefrom, LI does not teach every element of the claim.

The OA cites [0104], [0016] -- [0017], Fig. 8, [0057], [0040], and [0043] -- [0047] without explanation of the pertinence of these references other than to assert:

"1st", shows "verbatim entities" represented by "exact match."

With regard to Fig. 8, and

"stemming" reads on "grouping words by using word stemming""

with regard to [0057].

While LI [0016]-[0018] identifies the use of Latent Semantic Indexing (LSI) generally, the cited paragraph does not disclose a matrix representing the collection as claimed. The referenced paragraphs, nor any other section of LI, do not contain disclosure of stemming verbatim terms (those terms contained in the original documents before applying LSI) and using such stemmed terms in an LSI space by supplementing the verbatim terms. Importantly, LI teaches away from the use of LSI by including it in the Description of the Related Art and concluding that section with

... the past work has failed to address the problem of efficient processing of queries when they undergo query expansion or of reducing the size of the indices used to perform query expansion and processing.

LI [0104] discloses a generic hardware system for executing LI's invention, again without any reference to the use in the context of a reduced dimension LSI space.

LI [0057] mentions stemming, but only in the context of a step applied before using LI's invention, not used as claimed here. The exact match of LI Fig. 8 does not relate to *verbatim entities* as disclosed in the application. The application refers to the original terms of the reference documents as *verbatim*, see for example

*[0029] In some embodiments, the present invention includes a method for representing the latent semantic content of a plurality of documents, where each document contains a plurality of verbatim terms.*

LI [0040] discloses

replacing each word ... by its higher level semantic concept ... [T]he index table ... is obtained by combining words ... Thus, the row entries corresponding to "car", "auto", :automobile", and "sedan" appear as a single entry Sem1  
...

LI [0043]-[0047] discloses which types of words are amenable to being replaced (not stemmed and generalized) by a higher-level concept. The application discloses:

*[0049] A generalized entity is a semantic unit that is a short phrase of one or more stemmed words ... [0051] Generalized entities are created through extracting phrases between stop words, and stemming similar words. For example, the entity (retriev\* text\*) could be generated from any of the following phrases: retrieved textual, retrieving textual, or retrieved text.*

Clearly Sem1 is not an entity *generalized* from a *stemmed verbatim* entity as claimed. Car, auto and the other terms cannot be stemmed into Sem1 in any manner.

As a consequence, the invention as claimed in Claim 8, and Claims 10 – 13 dependent therefrom, are patentably distinct from LI.

Regarding Claim 9, LI does not teach every element of the claim.

The OA cites LI [0057], without explanation of the pertinence of this reference.

LI [0057] contains disclosure directed to estimating storage overhead. Such disclosure is totally irrelevant to *stemming ...* and *... generalizing ... ordered terms between stop words* and then replacing the original *verbatim* words with the stemmed and generalized words.

As a consequence, and in addition to the reasons cited regarding Claim 8, the invention as claimed in Claim 9 is patentably distinct from LI.

Regarding Claim 14, and Claims 15-17 dependent thereon, LI does not teach every element of the claim.

The OA cites LI [0016], [0040], [0035], and [0092], without explanation of the pertinence of these references.

LI [0016], while it identifies LSI generally, does not disclose any method of ranking results from a query directed to the LSI space. LI [0040] is disclosure directed to forming index tables, not to ranking the results of a query targeted at an LSI space, as claimed. LI [0035] discloses nothing about using the frequency of occurrence of a term as a means of ranking results from a query directed to an LSI space. LI [0092] discloses ranking candidate documents

based on degrees of relaxation; but discloses nothing about ranking results of a query to an LSI space based on the frequency of occurrence of terms.

As a consequence, the invention as claimed in Claim 14, and Claims 15-17 dependent thereon, is patentably distinct from LI.

Regarding Claim 18, and Claims 19-21 dependent thereon, LI does not teach every element of the claim.

The OA cites LI [0016], [0040], Fig. 8, [0057], [0040], [0048], [0038] -- [0040], [0060] -- [0062], and [0071] without explanation of the pertinence of these references other than to assert:

stemming" reads on "grouping words by using word stemming with regard to [0057].

Other than LI [0016], the paragraphs asserted as anticipating Claim 18 do not refer to LSI (an explicit limitation in each claim). With regard to LSI (and as noted above), LI teaches away from LSI in LI [0019]:

However, as noted above, the past work has failed to address the problem of efficient processing of queries ...

The OA cites LI [0038]-[0040] as anticipating the use of an LSI space to assess the closeness between a query and reference documents indexed into the space. These paragraphs of LI disclose no such approach. LI [0038]-[0040] disclose an unacceptable method of query expansion (not processing), and then additional index structures that are needed to practice LI's invention of query expansion.

Subsequently, LI [0041]-[0048] elaborate on the reduction in index size achievable when using LI's invention. The OA incorrectly asserts that use of multigranularity corresponds to the *generalized entity* found in the claims. Application [0051] states:

*Generalized entities are created through extracting phrases between stop words, and stemming similar words.*

LI teaches that multigranularity can be implemented by collecting words with similar semantic meaning into classes, e.g., LI Fig. 6 shows car, auto, automobile, sedan, garage, while LI Fig. 7 replaces those with the higher-level concept Sem1. Obviously this is not stemming as claimed.

As a consequence, the invention as claimed in Claim 18, and Claims 19-21 dependent thereon, are patentably distinct from LI.

Regarding Claim 22, LI does not teach every element of the claim.

The OA cites LI [0016], [0017], [0072], [0020], [0092], [0096], Fig. 2(b), [0015], [0097] -- [0099], and [0081] without explanation of the pertinence of these references.

While LI [0016] – [0017] identifies the use of Latent Semantic Indexing (LSI) generally, LI does not disclose a matrix representing the collection as claimed (in [0016] – [0017] or otherwise). None of the referenced paragraphs contain disclosure of *expansion phrases* indexed into an LSI space as claimed. Importantly, LI **teaches away** from the use of LSI by including it in the Description of the Related Art and concluding that section with

... the past work has failed to address the problem of efficient processing of queries when they undergo query expansion or of reducing the size of the indexes used to perform query expansion and processing.

No portion of LI teaches or suggests creation of an LSI space using *expansion phrases* as claimed. Further, LI's disclosure regarding query expansion does not include expansion phrases as disclosed, e.g., including replacing acronyms with full terms.

LI [0072] described query expansion according to LI's invention, not *phrase* expansion as claimed. LI [0020] is simply the summary of LI's query (not phrase) expansion invention. LI [0092] discloses ranking candidate documents based on degrees of relaxation; but discloses nothing about phrase expansion. [0096] – [0099] disclose a ranking graph that is clearly not LSI. While [0081] is disclosure regarding compression factor for storage – and thoroughly unrelated to the claim.

As a consequence, the invention as claimed in Claim 22 is patentably distinct from LI.

Regarding Claim 23, LI does not teach every element of the claim.

The OA cites LI [0035], [0092], [0096], [0097], Fig. 2(b), Fig. 2(a), [0089], [0016], and [0081] without explanation of the pertinence of these references other than the assertion:

"idiom" reads on "syntactically"

With regard to [0035] and [0092], and

"excluded" reads on "filtering out"

with regard to [0089] and [0035].

While LI [0016] identifies the use of Latent Semantic Indexing (LSI) generally, LI does not disclose a matrix representing the collection as claimed. The referenced paragraphs do not contain disclosure of the *expansion phrases* indexed into an LSI space as claimed. Importantly, LI **teaches away** from the use of LSI by including it in the Description of the Related Art and concluding that section with

... the past work has failed to address the problem of efficient processing of queries when they undergo query expansion or of reducing the size of the indexes used to perform query expansion and processing.

No portion of LI teaches or suggests creation of an LSI space using "expansion phrases" as claimed.

Further, LI's disclosure regarding query expansion does not include expansion phrases as disclosed, e.g., including replacing acronyms with full terms. LI's expansion phrases replace words with higher-level concepts. As claimed in Claim 23, embodiments of the present invention exclude idioms from the LSI space, not replace them with anything. In addition, *idiom* (a noun) cannot meaningfully be replaced with *syntactically* (an adverb).

As a consequence, the invention as claimed in Claim 23 is patentably distinct from LI.

Regarding Claim 24, LI does not teach every element of the claim.

The OA cites LI [0035], [0092], Claim 13, [0096], [0097], Fig. 2(b), Fig. 2(a), [0016], and [0081] without explanation of the pertinence of these references other than the assertion:

"idiom" reads on "syntactically"

With regard to [0035] and [0092], and

While LI [0016] identifies the use of Latent Semantic Indexing (LSI) generally, LI does not disclose a matrix representing the collection as claimed. The referenced paragraphs do not contain disclosure of the *expansion phrases* indexed into an LSI space as claimed. Importantly, LI **teaches away** from the use of LSI by including it in the Description of the Related Art and concluding that section with

... the past work has failed to address the problem of efficient processing of queries when they undergo query expansion or of reducing the size of the indexes used to perform query expansion and processing.

No portion of LI teaches or suggests creation of an LSI space using "expansion phrases" as claimed.

Further, LI's disclosure regarding query expansion does not include expansion phrases as disclosed, e.g., including replacing acronyms with full terms. LI's expansion phrases replace words with higher-level concepts. As claimed in Claim 23, embodiments of the present invention exclude idioms from the LSI space, not replace them with anything. In addition, *idiom* (a noun) cannot meaningfully be replaced with *syntactically* (an adverb).

As a consequence, the invention as claimed in Claim 23 is patentably distinct from LI.



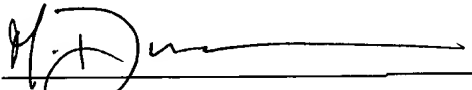
## CONCLUSION

With consideration of the above amendments to address objections and typographical errors along with remarks directed to rejections, the undersigned submits that this application is in condition for allowance, and such disposition is earnestly solicited. If the Examiner believes that the prosecution might be advanced by discussing the application with the undersigned, in person or over the telephone, we would welcome the opportunity to do so.

Respectfully submitted,

Date: March 4, 2004

KILPATRICK STOCKTON LLP  
607 14<sup>th</sup> Street, N.W.  
Suite 900  
Washington, DC 20005-2018  
(202) 508-5883

By:   
Michael J. Dimino  
Registration No. 44,657